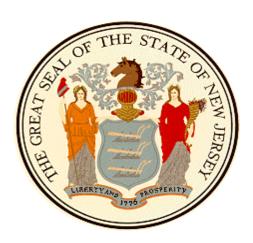
Estimating the Extent of Illicit Drug Abuse in New Jersey Using Capturerecapture Analysis



November 1998



Research & Information Systems Division of Addiction Services

Christine Todd Whitman Governor





Center for Substance Abuse Treatment SAMHSA

Produced under a contract funded by the
Center for Substance Abuse Treatment,
Substance Abuse and Mental Health Services Administration,
U.S. Department of Health and Human Services
Center for Substance Abuse Treatment
5600 Fishers Lane
Rockwall II, Suite 618
Rockville, Maryland 20857
301.443.5052

This report was supported by contract No. 270-96-0004 under the State Systems Development Program administered by the Division of State Programs, Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration.

Estimating the Extent of Illicit Drug Abuse in New Jersey Using Capture-recapture Analysis

Abate Mammo

Research and Information Systems
Division of Addiction Services
New Jersey Department of Health and Senior Services

November 1998

This report was supported by Contract No. 270-96-0004 under the State Systems Development Program administered by the Division of State Programs, Center for Substance Abuse treatment, Substance Abuse and Mental Health Services Administration.

Introduction

The ideal approach to treatment need estimation is the use of an objective quantitative assessment that allows for the diagnosis of the level of substance abuse or dependence. Such a diagnostic scheme (e.g., DSM-III-R), however, is limited by the ease of access to the study population and the will of this population to participate in often probing interviews that could easily be perceived to have legal consequences as well as stigma. Unlike for alcohol and cigarette use or abuse, such approaches usually fail to provide adequate data on the use or abuse of illicit drugs by householders.

The search for quantitative data that could be used for some comprehensive treatment needs assessment for the state and for its counties led us to apply the two sample capture-recapture model. We find that a substantial number of people abuse illicit drugs in the state and that the county distribution of abusers is consistent with our expectations.

Background

A method that was originally used for the study of salmon populations by Walton (1653) and formally introduced for the estimation of the size of populations using multiple independent samples taken over time by Peterson (1894) has now been qualified to provide estimates of similar populations that do not lend themselves to standard methods of estimation techniques (see Cormack (1968) for an extensive review of the statistics of this method). Later, the independence assumption was relaxed by providing methods of estimating dependencies (Bishop, Fienberg and Holland, 1975; Doscher and Woodward, 1983). We are concerned here with the estimation of

a closed population (i.e., change in the population occurs because of births or deaths) and follow the notation by Bishop et al. (1975). Models that entertain open populations are reviewed extensively elsewhere (Cormack, 1968).

The capture-recapture method assumes that there is a set of samples, lists, or rosters that contain subjects which are uniquely identified so that we can determine whether an individual was present or absent in any one of the samples. If there are k sources of data (i.e., samples) collected at non overlapping time points and not necessarily taken in any sequential order, it is assumed that the individual will fall into any one of the 2^k -1 cells and that the 2^k th cell is not observed in any of the k samples. The statistical problem is to estimate the size of the population which we know exists but has not been possible to sample.

To familiarize ourselves with the notations in such a model, we assume that there were two samples taken and the following cross-classifications were obtained.

		Second Sample (1998)				
First		Admitted	Not Admitted	Total		
Sample (1996)	Admitted	X ₁₁	X ₁₂	X ₁₊		
	Not admitted	\mathbf{X}_{21}	-			
	Total	X ₊₁				

 x_{ij} corresponds to the observed number of individuals in cell (i,j). The subscript 1 indicates that the individual is observed (admitted) in the sample and 2 indicates that it was not observed (not admitted) in the sample. For example, x_{21} represents the number of individuals who were

observed in the second sample (i.e., 1998) but not in sample one (i.e., 1996). Notice that x_{22} is the missing count in the cell designated by " -". Our objective is to estimate this missing count of individuals who were not reported in any of the two samples.

If we let n be the total number of individuals observed in the 2^k -1 cells (3 cells, in our example), then

$$n \cdot X_{11} \cdot X_{12} \cdot X_{21} \cdot \mathbf{j}_{i} \cdot \mathbf{j}_{j} \cdot X_{ij}$$
 (1)

The asterisk indicates that the observation in cell (2,2) is not included.

Let p_{11} be the probability of an individual being admitted in both 1996 and 1998, p_{1+} the probability of being admitted in 1996, p_{+1} the probability of being in 1998. If we assume that the two samples (or admissions in 1996 and 1998) are independent then $p_{11} = p_{1+} p_{+1}$, and if n is fixed then (x_{11}, x_{12}, x_{21}) has the multinomial distribution with probability function

$$\binom{n}{X_{11}, X_{12}, X_{21}} \frac{(p_{1\%}p_{\%1})^{X_{11}}[p_{1\%}(1\&p_{\%1})]^{X_{12}}[p_{\%1}(1\&p_{1\%})]^{X_{21}}}{[1\&(1\&p_{1\%})(1\&p_{\%1})]^n}$$
 (2)

The maximum likelihood estimates of the marginal probabilities, $p_{_{\!+\, 1}}$ and $p_{_{\!1+}}$ are given in Equation 3.

$$\hat{p}_{1\%}$$
 , $\frac{x_{11}}{x_{\%1}}$, $\hat{p}_{\%1}$, $\frac{x_{11}}{x_{1\%}}$ (3)

Suppose that n has a binomial distribution with sample size N and the probability of being

observed in at least one of the two samples, $p = 1-(1-p_{1+})(1-p_{+1})$. Then the probability of

$$\binom{N}{n}(p)^n(1\&p)^{N\&n} \tag{4}$$

selecting n out of N is

The maximum likelihood estimate (MLE) of N (N) is then given as the number of observed individuals, n divided by p. If we then substitute x_{11}/x_{+1} for \hat{p}_{1+} and x_{11}/x_{1+} for \hat{p}_{+1} into the MLE equation we get

$$\hat{N} : \frac{X_{1\%}X_{\%1}}{X_{11}} \tag{5}$$

Peterson (1894) derived the same formula to estimate the size of fish populations. Sekar and Deming (1949) considered situations where \mathbf{x}_{+1} and \mathbf{x}_{1+} were not fixed and showed that combining the multinomial probability function (Equation (2)) with the binomial probability function (equation (4)) will result in the same MLE estimators for N, \mathbf{p}_{1+} and \mathbf{p}_{+1} . Sekar and Deming (1949) also provided a formula (Equation (6)) for the asymptotic variance of \mathbb{N} (Bishop et al., 1975).

$$\widehat{Var}(\hat{N}) = \frac{X_{12}X_{21}X_{1\%}X_{\%1}}{X_{11}^3}$$
 (6)

In some situations, the presence or absence of an individual in the second sample may depend on whether or not that individual was present in the first sample. Naturally, this violates the independence assumption used earlier between the two samples. Models are available that correct biases that may originate from the lack of independence of samples¹.

For k samples, where there are 2^k incomplete cross-classifications with one missing cell, the estimation formulas can be generalized. Let $m(i)_{12...k}$ be the expected number of individuals in the (i_1,i_2,\ldots,i_k) cell of the 2^k table, where i_j $(j=1,2,\ldots,k)$ equals 1 if the individual is present in this cell or 2 if absent. Cell $(2,2,\ldots,2)$ is the missing cell so that $m_{22...2}=0$. Suppose also that the set S contains the 2^k -1 cells excluding the cell $(2,2,\ldots,2)$. Then n equals to $E^*x(i)_{12...k}$ where the summation runs over the set S. The probability that an individual falls in the cell (i_1,i_2,\ldots,i_k) is $m(i)_{12...k}/n$. If N is the total number of individuals in the population, N - n individuals are absent from the k samples. Following the 2-sample example we will proceed to estimate $m_{22...2}$ by $m^*_{22...2}$ as follows.

$$m_{22...2}^{(} - \frac{M_{odd}}{M_{even}}$$
 (7)

In equation (7), M_{odd} is the product of all $x(i)_{12...k}$ in S where the sum of the subscripts is equal to an odd number. M_{even} likewise refers to the product of those with even sums in their

$$\log m_{ii}$$
 ' $u \% u_1(i) \% u_2(j) \% u_{12}(ij)$

Where u = the grand mean of the logits of the expected cell counts

 $u_1(I)$ = main effect of variable 1 $u_2(j)$ = main effect of variable 2

 $u_{12}(ij)$ = two factor effect between variables 1 and 2

 $^{^{\}mathrm{1}}$ The log-linear model which assumes that all pair-wise relationships are present is given as

subscripts. The estimation of N and its asymptotic variance follow similar expressions as in the two sample model. The estimate for N is then given as follows:

$$\hat{N} = n + m^*_{22}$$

The capture-recapture model was applied to our Client Oriented Data Acquisition Process (CODAP) data by French (1977a, 1977b) to estimate the size of heroin abusers in New Jersey. Bonett, Woodward and Bentler (1986) recently suggested a linear model for the estimation of the size of a closed population using multiple recapture samples. Doscher and Woodward (1983) caution, however, that an attempt to estimate the size of a heterogenous population (different sampling probabilities) without stratification of the sample into homogenous groups would bias estimates arrived at using this method and suggest methods of correction for such problems. Following the advice by Doscher and Woodward (1983), Mammo (1995) estimated the numbers of heroin, cocaine and other drug abusers in New Jersey and for its 21 counties. This report closely follows his approach in 1995 except the minor exceptions.

Assumptions Used

We applied the two sample capture-recapture method to data obtained from the 1996 and 1998 Alcohol and Drug Abuse Data System (ADADS) to estimate the number of drug abusers in New Jersey. ADADS is the surveillance system maintained by our Division to monitor treatment activities for substance abuse and dependence in the state. We made the following assumptions:

1) An illicit drug abuser is one who was admitted for at least one illicit drug problem

- as primary, secondary or tertiary drug of choice. We excluded alcohol only admissions from the analysis.
- Admissions for a particular drug abuse treatment in a county in 1996 and 1998 were independent of each other (i.e., the fact that drug abusers were admitted in 1996 had no influence on their readmission probabilities in 1998).
- Substance abusers who sought treatment have characteristics similar to those who did not.
- 4) A heroin abuser is one who was admitted to treatment for a heroin or opiate abuse problem as the primary, secondary or tertiary drug of choice.
- A cocaine abuser is one who was admitted to treatment for a cocaine or crack problem as the primary, secondary or tertiary drug of choice after excluding heroin abusers and alcohol only abusers.
- 6) Other drug abusers are those who were admitted for treatment for drugs other than heroin, cocaine and crack.

We use 1996 and 1998 unduplicated admissions to minimize overlaps between the two samples. Almost all 1996 admissions were discharged from treatment before 1998 making them available for readmission (recapture) in 1998. Some of the differences we may find between the 1993 and 1998 estimates may in part be due to our use of more non overlapping samples in 1998 compared to 1993.

To improve homogeneity as advised by (Doscher & Woodward, 1983), separate estimates were made for heroin, cocaine, and other drugs within each county. By making separate estimates

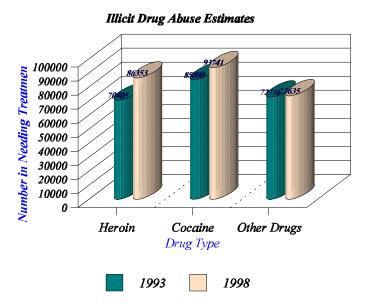
for each drug type within each of the 21 counties of New Jersey, we have tried to keep the drug abusing populations homogeneous. This allows us to assume equal probability of capture in the two samples for each drug type within a county thereby decreasing heterogeneity of populations.

Findings

Applying the capture-recapture method to treatment data provides useful estimates for the number of illicit drug abusers the majority of whom may end up in treatment². The validity of this method for the estimation of other substances such as alcohol is questionable. Household surveys are believed to produce more reliable estimates for alcohol abuse than for illicit drugs because of social desirability.

We estimate that there were 86,353 (95% C.I.: (84,441, 88,265)) heroin abusers in need of treatment in New Jersey up from 70,405 estimated in 1993. As expected, Essex county contributes the largest statewide estimated number of heroin abusers with 22,750 followed by Hudson with 8,703 and Union with 7,714 counties. The respective estimates in 1993 were 20,636, 8,757 and 6,247. Table 1 presents details of the 1998 estimates for each county for heroin, cocaine, and other drugs estimates.

Consistent with 1993, we estimate an even larger number of cocaine abusers (93,739 with a 95% C. I.: (89,077 to 98,402) in New Jersey compared to the number of heroin abusers. The number of cocaine abusers appears



² We expect most drug abusers to end up in treatment by themselves, through contact with the criminal justice system or through other factors influences.

to have gone up from the 1993 estimate of 85, 080 (95% C. I. :82,261, 87,899). Again, Essex county contributed the largest number of cocaine abusers with 10,427 (95% C. I. (8,648, 12,206)) to the statewide estimate. Somewhat unexpectedly, Monmouth follows Essex with 9,174 (95% C. I.: (7,889,10,459)) in its contributions. Camden with 8,109 (95% C. I. (6,655, 9,563)) and Hudson with 7,749 (95% C. I.: (6,387, 9,111)) also make significant contributions to the statewide estimate. Consistent with 1993 estimates, variations in heroin abuse by county remain higher than variations than cocaine abuse.

All other drugs combined contribute 73,635 (95% C. I. (67,600, 79,670)) of the 253,729 estimated illicit drug abusers in the state. We believe that the actual prevalence of illicit drug abuse in the state is much larger than the estimates suggest because of poly drug abuse.

In Table 2 we present the number of injection drug abusers estimated using the two-sample capture-recapture approach. We estimate that there are 26,975 injectors (95% C.I. (26,194,27,754) in the state almost all of whom are heroin injectors. As expected, Essex has the largest number of injectors with 4,336 (95% C. I.: (4,048, 4,624)) followed by Camden with 2,442 (95% C. I.: (2,160, 2,724)) and Hudson with 2,407 (95% C. I.: (2,161, 2,653)).

Table 3 presents 1993 and 1998 heroin, cocaine, and other drug estimates side by side for comparison purposes. The table shows that total drug abuse has increased by drug type with most of the increases occurring for heroin and cocaine.

Discussion

Our search for quantitative information that could be used for a more comprehensive treatment need assessment for the state and for its subregions prompted us to apply the two sample capture-recapture model. We find that a substantial number of people (253,729) are in need of treatment for their illicit drug abuse problems in the state. Of these, 86,353 need treatment for their heroin problems, 93,741 need treatment for their cocaine problems and 73,635 need treatment for other drug problems. Consistent with 1993 estimates, there are more cocaine abusers in the state than heroin abusers in 1998. While treatment need for heroin abuse and cocaine abuse appear to have gone up since 1993, need for treatment of other drug abuse appears to have stabilized at about 73,000. More importantly, injection drug use has also increased and the increase is consistent with the recent rise in injection drug use observed in the treatment population (Mammo, Schadl and Rodriguez, 1998).

Consistent with expectation, there is a substantial variation in heroin abuse by county with Essex county leading all counties in heroin and cocaine abuse and Monmouth county leading in other drug abuse.

The estimates presented here will supplement other studies in the Treatment Needs Assessment family of studies such as the 1998 Telephone Household Survey and the 1998 TANF Survey in assessing need and demand for treatment in the state and its counties.

References

- Bishop, Yvonne M.M., Fienberg, Stephen E. and Paul W. Holland. <u>Discrete Multivariate Analysis:</u>

 <u>Theory and Practice</u>. Cambridge, Massachusetts: The MIT Press, 1975.
- Bonett, Douglas G., Woodward, J. Arthur and P.M. Bentler. "A linear model for estimating the size of a closed population," British Journal of Mathematical and Statistical Psychology, 39:28-40, 1986.
- Cormack, R.M. "The statistics of capture-recapture methods," <u>Oceanographer Marine Biology Annual Review</u>, 6:455-506, 1968.
- Doscher, Mary-Lynn, and J. Arthur Woodward. "Estimating the size of sub-populations of heroin users: applications of log-linear models to capture/recapture sampling," <u>The International Journal of the Addictions</u>, 18(2):167-182, 1983.
- French, J.F. "Prevalence of heroin use in New Jersey," Unpublished report of Department of Health, Alcohol, Narcotic and Drug Abuse Unit, State of New Jersey, November, 1977(a).
- French, J.F. "Relative heroin prevalence," Unpublished report of Department of Health, Alcohol, Narcotic and Drug Abuse Unit, State of New Jersey, November 1977(b).
- Mammo, Abate, Richard Schadl and Gloria Rodriguez. "Recent Drug Abuse Variations in Newark,"

 Community Epidemiology Work Group, in Press, June 1999 Report.
- Mammo, Abate. <u>Estimating the Number of Drug Abusers in New Jersey Using a Capture-recapture</u>

 Model, A Report Submitted to CSAT, November 1, 1995.
- Peterson, C.G.J. Rrp. Dan. bio. Stn., 4:1-147, 1894.
- Sekar, C. C. and W. E. Deming. "On a method of estimating birth and death rates and the extent of registration" J. American Statist. Assoc. 44: (101-115, 1949.
- Walton, L. The Compleat Angler. Reprint Everyman's Library, London, 1653.

Table 1

Number of people who abuse illicit drugs in New Jersey, 1998

County	Heroin			Cocaine			Other Drugs			Total
		95% Cor	ifidence	95% Confidence			95% Confidence			
	Number	Number Limits		Number in	Limits		Number	Limits		
	Lower Upper			Lower Upper			Lower Upper			
	in Need			Need			in Need			
Atlantic	3,624	3,266	3,982	4,792	3,684	5,900	5,070	3,091	7,049	13,486
Bergen	4,511	3,858	5,164	5,479	4,605	6,353	4,426	3,069	5,783	14,416
Burlington	1,846	1,457	2,235	4,420	3,181	5,659	4,437	1,818	7,056	10,703
Camden	6,530	5,868	7,192	8,109	6,655	9,563	5,225	3,558	6,892	19,864
Cape May	830	599	1,061	1,174	762	1,586	1,895	1,170	2,620	3,899
Cumberland	1,012	830	1,194	2,506	1,659	3,353	1,554	616	2,492	5,072
Essex	22,750	21,814	23,686	10,427	8,648	12,206	6,968	4,221	9,715	40,145
Gloucester	1,648	1,280	2,016	3,050	2,067	4,033	2,204	1,354	3,054	6,902
Hudson	8,703	8,084	9,322	7,749	6,387	9,111	5,662	3,683	7,641	22,114
Hunterdon	656	363	949	641	272	1,010	1,362	545	2,179	2,659
Mercer	2,204	1,950	2,458	7,081	6,012	8,150	3,878	2,300	5,456	13,163
Middlesex	6,289	5,764	6,814	6,181	4,826	7,536	4,231	3,063	5,399	16,701
Monmouth	3,796	3,495	4,097	9,174	7,889	10,459	7,399	5,497	9,301	20,369
Morris	2,656	2,252	3,060	2,282	1,632	2,932	1,799	1,113	2,485	6,737
Ocean	2,955	2,606	3,304	3,746	2,934	4,558	4,130	2,950	5,310	10,831
Passaic	6,102	5,591	6,613	6,182	4,867	7,497	4,593	2,230	6,956	16,877
Salem	186	122	250	1,316	412	2,220	663	184	1,142	2,165
Somerset	1,356	1,082	1,630	1,961	1,022	2,900	1,873	797	2,949	5,190
Sussex	573	422	724	702	269	1,135	1,709	118	3,300	2,984
Union	7,714	7,183	8,245	5,487	4,080	6,894	3,668	1,873	5,463	16,869
Warren	412	277	547	1,282	595	1,969	889	558	1,220	2,583
New Jersey	86,353	84,441	88,265	93,741	89,078	98,404	73,635	67,600	79,670	253,729

Source: Drug need estimates are made using data from the Alcohol and Drug Abuse Data System (ADADS) of 1996 and 1998 and applying a two-sample capture-recapture estimation approach.

Note: Heroin estimates refer to any heroin admission as primary, secondary or tertiary drug of choice at the time of treatment. Cocaine estimates refer to any cocaine admissions as primary, secondary or tertiary drug of choice at the time of admission. Other drug estimates refer to the residual of drug abusers/dependents after heroin, cocaine and alcohol only are excluded. Both heroin and cocaine estimates are made after alcohol only admissions are excluded.

 $\label{eq:Table 2} \textbf{Estimated number of injection drug abusers in New Jersey, 1998}$

		1998					
County	Number of	95% Co	Number of Injectors in				
	Injectors	Lower	Upper	1993			
Atlantic	1,914	1,707	2,121	1,914			
Bergen	1,726	1,425	2,027	1,454			
Burlington	695	544	846	703			
Camden	2,442	2,160	2,724	2,281			
Cape May	577	389	765	391			
Cumberland	536	431	641	569			
Essex	4,336	4,048	4,624	4,919			
Gloucester	701	551	852	747			
Hudson	2,407	2,161	2,653	2,388			
Hunterdon	244	115	373	103			
Mercer	1,293	1,139	1,447	1,430			
Middlesex	1,967	1,769	2,165	1,513			
Monmouth	1,543	1,399	1,687	1,516			
Morris	610	497	723	495			
Ocean	1,237	1,071	1,403	714			
Passaic	2,022	1,800	2,244	1,464			
Salem	100	40	160	110			
Somerset	381	287	475	410			
Sussex	200	123	277	131			
Union	1,770	1,608	1,932	1,581			
Warren	273	116	430	157			
New Jersey	26,975	26,194	27,754	24,989			

Table 3

Number of people who abuse illicit drugs in New Jersey, 1998

		1993				1998				
County	Heroin	Cocaine	Other Drugs	Total	Heroin	Cocaine	Other Drugs	Total		
Atlantic	3,066	3,625	3,095	9,786	3,624	4,792	5,070	13,486		
Bergen	2,843	5,573	4,564	12,980	4,511	5,479	4,426	14,416		
Burlington	1,023	3,682	2,748	7,453	1,846	4,420	4,437	10,703		
Camden	5,475	7,617	3,379	16,470	6,530	8,109	5,225	19,864		
Cape May	512	1,094	1,407	3,014	830	1,174	1,895	3,899		
Cumberland	956	2,421	1,894	5,271	1,012	2,506	1,554	5,072		
Essex	20,636	14,070	7,811	42,516	22,750	10,427	6,968	40,145		
Gloucester	983	2,563	3,016	6,562	1,648	3,050	2,204	6,902		
Hudson	8,757	5,951	3,506	18,214	8,703	7,749	5,662	22,114		
Hunterdon	227	426	1,165	1,817	656	641	1,362	2,659		
Mercer	2,133	5,500	5,243	12,875	2,204	7,081	3,878	13,163		
Middlesex	5,281	6,087	5,876	17,244	6,289	6,181	4,231	16,701		
Monmouth	3,334	6,391	5,600	15,326	3,796	9,174	7,399	20,369		
Morris	1,362	2,723	3,993	8,078	2,656	2,282	1,799	6,737		
Ocean	1,175	2,369	5,147	8,691	2,955	3,746	4,130	10,831		
Passaic	4,455	4,513	3,433	12,401	6,102	6,182	4,593	16,877		
Salem	198	772	805	1,775	186	1,316	663	2,165		
Somerset	995	1,474	3,709	6,177	1,356	1,961	1,873	5,190		
Sussex	527	640	1,138	2,306	573	702	1,709	2,984		
Union	6,247	7,073	4,255	17,575	7,714	5,487	3,668	16,869		
Warren	220	515	933	1,668	412	1,282	889	2,583		
New Jersey	70,405	85,080	72,716	228,201	86,353	93,741	73,635	253,729		

